Journal of Education and Practice
ISSN 2222-1735 (Paper) ISSN 2222-288X (Online)
Vol.16, No.13, 2025

Enhancing Learners' Deep Writing Performance in Generative AI-Assisted Environments through Self-Assessment Prompts

Yiheng Lou [1], Binghang Yu [1], Xinjie Chen[1*](Corresponding author)
1. Zhejiang Key Laboratory of Intelligent Education Technology and Application, Zhejiang Normal University, Jinhua, 321004, Zhejiang, China
* E-mail of the corresponding author: chenxinjie217@gmail.com

The research is financed by Zhejiang Provincial Science and Technology Innovation Program (New Young Talent Program) for College Students No. 2025R404AB062(Sponsoring information)

Abstract

As generative artificial intelligence (GenAI) rapidly permeates educational practice, its influence on students' writing behaviors has drawn increasing scholarly attention. While GenAI can substantially reduce the cognitive burden of writing and enhance textual quality, concerns have emerged regarding its potential to induce cognitive offloading, thereby weakening learners' higher-order thinking and reflective engagement. Addressing this issue requires pedagogical interventions capable of sustaining learners' active cognitive participation during human–AI collaborative writing. To address this challenge, this study investigates whether introducing self-assessment prompts, as a form of metacognitive intervention, can effectively enhance learners' deep writing performance in GenAI-assisted environments. Employing a randomized pre-test-post-test control group design, the study assigned 62 university students to an experimental group (receiving self-assessment prompts) and a control group (no prompts). An analysis of data from the Deep Writing Process Scale using a mixed-design ANOVA revealed that: 1) a statistically significant interaction effect between time and group was found on the total scale score ($p = .046$), indicating that the self-assessment intervention significantly promoted the learners' overall deep writing process; 2) among the sub-dimensions, the intervention's positive impact was particularly significant on "Effective Communication" ($p = .011$), with a marginally significant positive trend for "Learning Perseverance" ($p = .094$); and 3) the effects on "Learning to Learn" and "Self-Directed Learning Tendency" were not significant. This research confirms that the enabling effects of technology are highly contingent upon effective pedagogical design. By functioning as a metacognitive scaffold, simple self-assessment prompts can effectively counteract cognitive inertia, ensuring learners maintain their core role as active, reflective agents in the human-AI collaborative writing process.

Keywords: Generative AI, EFL writing, Deep Writing, Self-Assessment

DOI: 10.7176/JEP/16-13-05

Publication date: December 30th 2025

1. Introduction

The emergence of Generative AI (GenAI) is profoundly reshaping the educational landscape, demonstrating transformative potential, particularly in the domain of writing pedagogy (Dong, 2025). By assisting learners with the entire writing process—from ideation and drafting to revision—GenAI significantly lowers the technical barriers to writing and opens up new avenues for enhancing textual quality. A consensus is emerging within the scholarly community that the judicious use of this powerful tool could catalyze a pivotal shift in writing instruction: from a traditional, product-oriented assessment to a more process-oriented approach that emphasizes the development of learners' cognitive and academic abilities (Shen, 2024).

However, this unprecedented convenience has also given rise to profound concerns, with the principal risk being the potential for "Cognitive Offloading" among learners (Chen, 2025). A growing body of research indicates that when cognitively demanding tasks—such as deep thinking, knowledge construction, and the refinement of arguments—can be effortlessly delegated to AI, learners may develop an over-reliance on the technology, thereby diminishing their capacity for proactive thought and critical analysis. In such a scenario, AI can readily devolve from a "cognitive partner" that inspires creativity into an "intelligent ghostwriter" that supplants critical thought. This creates a fundamental conflict with the core objective of deep writing, which is to cultivate higher-order thinking skills.

Current research on GenAI-assisted writing has largely concentrated on its impact on the writing product (e.g., textual fluency) or focused on the technical optimization of prompt engineering (Liu, 2024). A critical research gap remains, however: how can we design pedagogical interventions that guide learners to engage in deep

metacognitive reflection while using AI? To address the aforementioned issue of cognitive offloading, self-assessment, a core strategy in self-regulated learning, presents a promising solution. By prompting learners to monitor and evaluate their work against established criteria, it encourages them to become more active and conscious participants in their learning. Although theoretically promising, the efficacy of systematically integrating this strategy into the human-AI collaborative writing process awaits empirical validation.

In light of this, the present study aims to investigate, through a controlled experiment, whether the introduction of external self-assessment prompts can effectively enhance learners' deep writing processes in a GenAI-assisted environment. We posit the following core hypothesis: compared to a control group using only standard GenAI, the experimental group receiving additional self-assessment prompts will achieve significantly higher scores on the Deep Writing Process Scale. This research seeks to provide insights of both theoretical and practical value for designing effective pedagogical interventions in the age of AI to foster learners' deep writing capabilities.

2. Literature review

2.1 Deep Writing

Deep writing is a higher-order cognitive practice that conceptualizes writing as a process for knowledge construction, critical analysis, and reflective practice (Chen, 2022). In contrast to "Surface Writing," which focuses merely on information retelling, deep writing requires authors to transition from being "transporters" of knowledge to "constructors" of meaning. This process necessitates that learners integrate and evaluate information, amalgamate it with their personal knowledge systems, and ultimately produce a personal perspective that is both original and logically rigorous.

Drawing upon previous research in the field, this study operationalizes deep writing into four interconnected core dimensions: Effective Communication, Learning to Learn, Learning Perseverance, and Self-Directed Learning Tendency. Together, these four dimensions constitute the internal requirements and external manifestations of deep writing.

Effective Communication: This serves as the objective of deep writing. It demands not only clarity and accuracy in language but also emphasizes the profundity and persuasiveness of the ideas conveyed. A deep writer must construct a logically coherent argument to achieve insightful and impactful communication (Wang, 2012).

Learning to Learn: This constitutes the metacognitive core of deep writing. Deep writing is, in essence, a metacognitive activity, requiring learners to engage in continuous planning, monitoring, and reflection throughout the entire writing process. This process of self-regulation is the central embodiment of the "learning to learn" competency (Barnard, 2009).

Learning Perseverance: This provides the emotional and volitional support for deep writing. When faced with the challenges of refining arguments and organizing logical structures, learners must possess the volitional quality of persistence and a commitment to continuously polishing their work. This is an indispensable psychological trait for the completion of deep writing (Downer, 2007).

Self-Directed Learning Tendency: This is the motivational source of deep writing. Deep writing requires learners to proactively set goals, seek information, and take ownership of the final output. This process significantly strengthens the learners' role transformation from passive recipients of knowledge to active inquirers.

In summary, this study defines deep writing as a multidimensional construct composed of the four aforementioned dimensions. Consequently, the Deep Writing Process Scale was selected as the core measurement instrument to assess the effectiveness of the intervention in this research.

2.2 Generative AI

Generative AI is reshaping the ecosystem of writing pedagogy as a powerful form of cognitive scaffolding. It can support learners throughout the entire writing process, from brainstorming during the ideation phase and framework construction in the drafting stage to real-time polishing during revision. These functionalities substantially lower the technical threshold for writing, enhance efficiency, and are particularly effective in helping learners overcome writing anxiety and initiate the writing process (Fadilah, 2023).

However, the uncritical use of this "scaffolding" can also present grave challenges, the core risk of which is inducing "Cognitive Offloading" in learners. When cognitively demanding tasks such as thinking, organizing, and argumentation can be effortlessly delegated to AI, the tool may transform from an assistive instrument into a

cognitive substitute, directly undermining the core processes emphasized in deep writing. This manifests in several ways:

First is the blunting of critical thinking. Learners may uncritically accept AI-generated content, overlooking the need to scrutinize its factual accuracy, logical coherence, or even latent biases. Consequently, writing degenerates from an active, critical process into one of passive reception. This is followed by an erosion of learner agency. Over-reliance on AI can lead to the displacement of a student's personal voice and unique perspective by the machine's formulaic expressions, rendering the text devoid of substantive thought. Finally, there is an absence of metacognitive monitoring. The fluency of AI's output can bypass the indispensable stages of internal reflection inherent in traditional writing (e.g., "Is my argument consistent?"), thereby hindering the development of learners' metacognitive abilities (Yan, 2024).

Therefore, the core challenge confronting writing education today is no longer a question of technological "availability," but of pedagogical "guidance." The objective of education must be to reap the benefits of technological convenience while mitigating its potential to erode the deep learning process. This necessitates the exploration of new pedagogical strategies that empower learners to become critical users of AI tools, rather than passive recipients, ensuring that technology always serves—rather than supplants—human growth.

2.3 Self-Assessment

Self-assessment is not merely an evaluative activity but a core metacognitive strategy. It requires learners to systematically scrutinize and judge their own learning processes and outcomes against explicit standards, such as rubrics or goals (Yan, 2017). Within the theoretical framework of Self-Regulated Learning (SRL), self-assessment constitutes the cornerstone for learners as active constructors of their knowledge. In the dynamic cycle of planning, monitoring, and reflection that characterizes SRL, self-assessment primarily drives the monitoring and reflection phases: it helps learners to identify the "gap between their current performance and their goals" and thereby triggers corresponding strategic adjustments. In essence, self-assessment is the key mechanism through which learners transition from passive recipients of information to active managers of their own learning.

In the specific context of GenAI-assisted writing, the role of self-assessment is particularly crucial. This study conceptualizes it as an effective cognitive intervention capable of counteracting "cognitive offloading" and promoting deep learning. Its mechanism of action manifests in the following three ways:

First, it activates metacognitive monitoring (Morphew, 2021). Unguided use of AI often leads to a fluent but superficial "autopilot" mode of engagement. External self-assessment prompts act as a "cognitive speed bump" and a form of scaffolding, compelling learners to temporarily disengage from their interaction with the AI and shift their perspective from that of a "user" to that of an "evaluator." This, in turn, reactivates critical appraisal of the text's logic and quality.

Second, it reshapes learner agency and responsibility. Self-assessment underscores the centrality of human cognition in human-AI collaboration. When learners are required to evaluate an AI-generated draft against set criteria, they are endowed with the final authorial judgment. This process compels learners to assume academic responsibility for the quality of their work, thus establishing their agency as the author.

Finally, it promotes strategic reflection (Wilby, 2022). Effective self-assessment focuses not only on the writing "product" but also on the "process" of human-AI interaction. By guiding students to reflect on questions such as, "What strategies did I use to optimize the AI's output?", self-assessment can foster the development of more efficient and critical modes of AI collaboration, leading to the true actualization of "learning to learn."

In conclusion, this study does not treat self-assessment as an ancillary step to be completed after writing, but rather positions it as a key transformative mechanism deeply integrated into the process of human-AI interaction. We argue that self-assessment, guided by external prompts, can elevate the human-AI interaction from a passive, task-driven relationship to an active, reflective learning partnership, thereby creating the necessary cognitive conditions for deep writing to occur.

3. Methodology

3.1 Research Design

This study employed a randomized pre-test-post-test control group design to examine the causal effect of self-assessment prompts on learners' deep writing processes within a GenAI-assisted writing environment.

The independent variable was the type of intervention, which consisted of two levels: an experimental group (n = 31) and a control group (n = 31). Both groups used the same large language model to complete the writing tasks. The sole distinction was that the experimental group received additional structured, system-generated prompts designed to guide their self-assessment, whereas the control group did not receive any such prompts.

The dependent variable was the learners' deep writing process performance, measured by the Deep Writing Process Scale. This scale comprises four sub-dimensions—"Effective Communication," "Learning to Learn," "Learning Perseverance," and "Self-Directed Learning Tendency"—as well as a total score.

Through random assignment, this study ensured the statistical comparability of the two groups on various initial characteristics (e.g., baseline writing proficiency, motivation) prior to the intervention. The pre-test-post-test design allows for a precise examination of whether there are significant differences in the changes observed in the deep writing process performance between the two groups, thereby enabling a more robust inference of the net effect of the intervention.

3.2 Participants

A total of 62 undergraduate students participated in this study. They were recruited through campus announcements and volunteered to take part in the experiment. All participants provided informed consent before the formal commencement of the study and were explicitly informed that their data would be used solely for academic research purposes and that their personal privacy would be strictly protected.

To achieve the research objectives, all participants were randomly assigned to either the experimental group (n = 31) or the control group (n = 31).

Prior to the experimental intervention, to verify the homogeneity of the two groups, the study collected and compared their basic demographic information and initial abilities related to the research variables. Statistical analysis revealed no significant differences between the experimental and control groups in terms of gender composition, age, pre-test writing scores, or initial self-assessment capabilities (all $p > .05$). This result indicates that the random assignment was effective and that the two groups were at a comparable baseline before the intervention began, thus providing a reliable foundation for the subsequent accurate assessment of the intervention's net effect.

3.3 Measurement Instruments and Materials

3.3.1 Deep Writing Process Scale (DWPS)

The core dependent variable in this study was measured using a scale adapted from the Deep Writing Process Scale by Chen (2022). The scale consists of 12 items rated on a 5-point Likert scale (1 = completely disagree, 5 = completely agree). It assesses the degree of learners' engagement in the writing process across the following four dimensions, with each dimension comprising 3 items:

Effective Communication: This dimension assesses learners' willingness and ability to express and exchange ideas. (Sample item: "I am able to evaluate my classmates' essays and provide reasonable and valuable suggestions.")

Learning to Learn: This dimension focuses on learners' metacognitive reflection and constructive engagement. (Sample item: "During writing classes, I seriously reflect on every viewpoint that is presented.")

Learning Perseverance: This dimension measures learners' resilience and persistence when faced with writing challenges. (Sample item: "Even if an essay takes a long time to complete, I will persist in finishing it.")

Self-Directed Learning Tendency: This dimension evaluates learners' proactivity and goal-oriented behavior in their studies. (Sample item: "I set goals for myself in order to achieve better grades on my essays.")

The scale has demonstrated good validity and reliability in previous research. For the sample in the present study, the Cronbach's alpha for the overall scale was .847, and the reliability coefficients for the sub-dimensions ranged from .692 to .727, indicating a high degree of reliability. Scores were calculated by averaging the items for each dimension and for the total scale. Higher scores represent a better performance in the deep learning process.

3.3.2 Writing Task and AI Tool

All participants were required to complete a standardized argumentative essay. The topic was selected from past CET-6 (College English Test Band 6) examination prompts, such as "The Impact of Artificial Intelligence on the

Future of Education." Both groups used the same version of a large language model (Gemini 2.5 Pro) as their writing assistance tool.

3.3.3 Self-Assessment Prompts

This was the core intervention material for the study, presented exclusively to the experimental group. At three key junctures in the writing process (after completing the outline, the first draft, and the revision), the system automatically presented structured evaluation prompts to the participants in the experimental group. The content of these prompts included:

Logical Coherence Assessment: "Please compare your current draft with your outline. Is the logic between the paragraphs coherent? Please rate it on a scale of 1-5."

Argument Quality Assessment: "Do you believe your core argument is supported by sufficient evidence? Please rate it on a scale of 1-5 and provide your reasoning."

Next-Step Action Plan: "Based on the evaluation above, please briefly describe your revision plan for the next step."

3.4 Experimental Procedure

The experimental procedure of this study was conducted over a period of eight weeks within a single academic semester and was divided into four distinct phases.

First, during the preparation and pre-test phase (Weeks 1-2), the research team recruited sixty university students to serve as participants through campus channels. After signing informed consent forms, all participants completed a series of pre-test questionnaires to collect baseline data for examining the homogeneity between the groups. Subsequently, they were randomly assigned to either the experimental group (n = 31) or the control group (n = 31).

Next, during the training and familiarization phase (Weeks 3-4), all participants attended a standardized training session on the fundamentals of English writing. This was done to control for pre-existing differences in knowledge. Following this, participants in both groups logged into and familiarized themselves with their designated AI writing environments.

The core experimental intervention phase lasted for three weeks (Weeks 5-7). Throughout this period, all participants were required to complete one writing task per week using their designated AI agent. The control group used the standard version of the large language model for writing assistance. In contrast, the experimental group used a version that integrated the self-assessment feature. This version automatically displayed prompts at key junctures in the writing process to guide participants in conducting structured reflection and evaluation of their own writing.

Finally, during the post-test phase (Week 8), all participants completed the core measurement instrument of this study, the Deep Writing Process Scale. Upon completion of data collection, the research team conducted a debriefing session for all participants to explain the study, and expressed their gratitude for the participants' involvement.

4. Results

This study employed a series of 2 (Group: experimental, control) × 2 (Time: pre-test, post-test) mixed-design analyses of variance (ANOVAs) to examine the effect of the intervention on the total score and the four sub-dimensions of the Deep Writing Process Scale.

4.1 Descriptive Statistics

Table 1 presents the means and standard deviations for both groups on all dependent variables at pre-test and post-test. The data indicate that the means for all measures were comparable between the two groups at the pre-test. Following the intervention, the post-test means for the experimental group were higher than those of the control group on the total scale score and across all four sub-dimensions.

Table 1. Pre-test and Post-test Means and Standard Deviations for the Total Score and Sub-dimensions of the Deep Writing Process Scale for Both Groups

Measurement Variable	Group	Pre-test M (SD)	Post-test M (SD)
Effective Communication	Experimental	3.25(.60)	3.93(.59)
	Control	3.54(.61)	3.84(.66)
Learning to Learn	Experimental	3.62(.59)	4.02(.58)
	Control	3.56(.61)	3.92(.67)
Learning Perseverance	Experimental	3.70(.53)	4.17(.49)
	Control	3.80(.65)	3.98(.69)
Self-Directed Learning Tendency	Experimental	3.80(.40)	4.12(.45)
	Control	3.74(.46)	3.92(.58)
Total Score	Experimental	3.67(.37)	4.06(.45)
	Control	3.70(.36)	3.92(.53)

Note: M = Mean; SD = Standard Deviation.

4.2 Inferential Statistics

A mixed-design ANOVA conducted on the total scale score revealed a statistically significant interaction effect between time and group, $F(1, 60) = 4.155$, $p = .046$, $\eta p^2 = .065$. This result indicates that the pattern of change in the deep writing process scores over time was significantly different between the two groups.

To further investigate the specific mechanisms of the intervention, separate mixed-design ANOVAs were conducted for each of the four sub-dimensions:

4.2.1 Effective Communication

The analysis revealed a significant main effect of time, $F(1, 60) = 44.86$, $p < .001$, $\eta p^2 = .43$. More importantly, the interaction effect between time and group was also significant, $F(1, 60) = 6.87$, $p = .011$, $\eta p^2 = .10$. This indicates that the magnitude of improvement on the "Effective Communication" dimension was significantly greater for the experimental group than for the control group.

4.2.2 Learning to Learn

The analysis found no significant interaction effect between time and group, $F(1, 60) = 0.051$, $p = .822$, $\eta p^2 = .001$. However, there was a significant main effect of time, $F(1, 60) = 28.437$, $p < .001$, $\eta p^2 = .322$, indicating that participants in both groups showed significant improvement on this dimension.

4.2.3 Learning Perseverance

The analysis showed a marginally significant interaction effect between time and group, $F(1, 60) = 2.901$, $p = .094$, $\eta p^2 = .046$. This suggests a trend that the improvement on the "Learning Perseverance" dimension might be greater for the experimental group than for the control group. The main effect of time was also significant, $F(1, 60) = 14.805$, $p < .001$, $\eta p^2 = .198$.

4.2.4 Self-Directed Learning Tendency

The analysis found no significant interaction effect between time and group, $F(1, 60) = 0.905$, $p = .345$, $\eta p^2 = .015$. However, the main effect of time was significant, $F(1, 60) = 11.834$, $p < .001$, $\eta p^2 = .165$, indicating that the scores of participants in both groups increased significantly over time on this dimension.

In summary, the results of the statistical analyses provide partial support for the research hypothesis. The data show that the intervention incorporating self-assessment had a significant positive effect on the learners' overall

deep writing process performance (p = .046), particularly on the "Effective Communication" dimension (p = .011). Furthermore, a positive trend of change was observed for the "Learning Perseverance" dimension (p = .094). However, no significantly greater improvement was found for the experimental group compared to the control group on the "Learning to Learn" and "Self-Directed Learning Tendency" dimensions.

5. Discussion

This study aimed to investigate whether the introduction of self-assessment prompts could effectively enhance learners' deep writing process performance in a GenAI-assisted writing environment. The findings offer robust, yet nuanced, support for our central hypothesis. Overall, the intervention that integrated self-assessment prompts significantly facilitated the learners' overall deep writing process; however, its impact varied across the different dimensions.

The central finding of this study is that, compared to the control group, the experimental group that received self-assessment prompts exhibited a significantly greater magnitude of improvement on the total score of the Deep Writing Process Scale. This result indicates that simple external prompts serve as an effective cognitive intervention, capable of counteracting the pervasive risk of "cognitive offloading" in AI-assisted writing. This finding aligns closely with the principles of Self-Regulated Learning (SRL) theory: in a standard AI-writing workflow, learners can easily fall into a passive role of reception and integration, thereby bypassing the critical "monitoring" and "reflection" stages of the SRL cycle. In contrast, the self-assessment prompts in this study acted as a "cognitive interrupter," which compelled learners to pause and scrutinize their work, thereby effectively "reactivating" the otherwise neglected metacognitive cycle and ultimately enhancing the overall deep learning process (Andrade, 2019).

Among the specific dimensions, the positive effect of the intervention was most pronounced for "Effective Communication." We attribute this to the fact that the act of self-assessment is, in essence, a form of deep "self-communication" (Muflihah, 2022). When required to evaluate the logic and argumentation of their text, learners are compelled to switch from the immersive perspective of the "author" to the critical perspective of the "reader." This act of perspective-taking prompts them to think more deeply about how to make their points more persuasive, which is the core of effective communication. While AI can generate language, it is the self-assessment process that guides learners to assume cognitive ownership over the ultimate communicative effectiveness of that language (Li, 2025).

Intriguingly, the intervention's effect on the "Learning Perseverance" dimension showed a marginally significant positive trend. This is consistent with the findings of Zhang (2023). We posit that this may be because self-assessment decomposes a large, ambiguous task ("write a good essay") into a series of concrete, controllable micro-evaluative tasks ("is the argument in this paragraph sufficient?"). This task decomposition can reduce learners' cognitive load and feelings of being overwhelmed. As learners identify and resolve these specific issues through self-assessment, the resulting micro-experiences of success accumulate, thereby enhancing their self-efficacy, which in turn transforms into greater learning perseverance (Zhu, 2024). The fact that this effect did not reach full statistical significance may be attributable to the relatively short duration of the intervention, as the cultivation of a relatively stable psychological trait such as learning perseverance likely requires a more extended period of engagement.

However, this study did not find that the intervention led to significantly greater gains for the experimental group in the dimensions of "Learning to Learn" and "Self-Directed Learning Tendency." A plausible explanation for this is that these two dimensions reflect deeper-seated learning habits and personal dispositions. A three-week intervention driven by external prompts may be sufficient to guide learners in executing the act of self-assessment, but it is likely insufficient for them to internalize this behavior as a spontaneous and sustainable learning strategy. As the results showed, participants in both groups improved significantly on these two dimensions (a significant main effect of time), suggesting that the act of using AI for writing practice is, in itself, a beneficial learning activity (Kim, 2025). However, making the leap from "behavioral compliance" to "capability internalization" likely requires a more prolonged and varied pedagogical design, one that goes beyond mere external prompts.

6. Conclusion

6.1 Conclusion and Implications

This study demonstrates through a randomized controlled experiment that, in the context of GenAI-assisted

writing, the introduction of simple self-assessment prompts as a metacognitive scaffold can significantly facilitate the overall deep writing process among university students. Specifically, the intervention was particularly effective in enhancing learners' "Effective Communication" competency and demonstrated a positive trend in fostering their "Learning Perseverance."

The core conclusion of this research is that the enabling effects of technology do not occur automatically but are highly contingent upon effective pedagogical design. In increasingly intelligent learning environments, the role of the educator should not be that of a mere technology provider, but rather that of a designer and facilitator of the learning process. Our findings provide clear empirical evidence and viable practical pathways for effectively designing human-AI collaborative learning environments in the age of AI to counteract potential cognitive inertia and cultivate learners' higher-order competencies.

6.2 Limitations and Future Directions

Although this study has yielded valuable findings, it is not without several limitations. First, the sample's representativeness is limited (N = 62, from a single university), which constrains the generalizability of the findings. Second, the short duration of the intervention (three weeks) may have been insufficient to produce a profound impact on cognitive traits, such as "Learning to Learn," that require long-term internalization. Finally, the mode of measurement was singular, relying primarily on a self-report scale. Future research should incorporate more objective, process-oriented data.

Future research could further validate the generalizability of this intervention by extending it to more diverse learner groups and disciplinary contexts. It would also be worthwhile to compare different types, frequencies, and levels of personalization in self-assessment prompts to identify more effective intervention designs. In addition, multimodal methods such as eye-tracking or physiological sensing may help uncover the cognitive and affective mechanisms through which self-assessment shapes learners' engagement in AI-assisted writing.

References

Andrade, H. L. (2019). A critical review of research on student self-assessment. Frontiers in Education, 4, 87. https://doi.org/10.3389/feduc.2019.00087

Barnard, L., Lan, W. Y., To, Y. M., Paton, V. O., & Lai, S.-L. (2009). Measuring self-regulation in online and blended learning environments. The Internet and Higher Education, 12(1), 1 ～ 6. https://doi.org/10.1016/j.iheduc.2008.10.005

Chen, A., Xiang, M., Zhou, J., Jia, J., Shang, J., Li, X., Gašević, D., & Fan, Y. (2025). Unpacking help-seeking process through multimodal learning analytics: A comparative study of ChatGPT vs Human expert. Computers & Education, 226, 105198. https://doi.org/10.1016/j.compedu.2024.105198

Chen, Y., Li, M., Huang, C., Han, Z., Hwang, G., & Yang, G. (2022). Promoting deep writing with immersive technologies: An SVVR -supported Chinese composition writing approach for primary schools. British Journal of Educational Technology, 53(6), 2071～2091. https://doi.org/10.1111/bjet.13247

Dong, J., & Wang, H. (2025). Exploring the effectiveness of a DDL-GenAI teaching mode in students' EFL writing classroom. Computer Assisted Language Learning, 1-50. https://doi.org/10.1080/09588221.2025.2577369

Downer, J. T., Rimm-Kaufman, S. E., & Pianta, R. C. (2007). How do classroom conditions and children's risk for school problems contribute to children's behavioral engagement in learning? School Psychology Review, 36(3), 413～432. https://doi.org/10.1080/02796015.2007.12087938

Fadilah, N. (2023). The impact of visual media on enhancing students' comprehension of islamic religious education lessons. WARAQAT : Jurnal Ilmu-Ilmu Keislaman, 8(2), 203 ～ 210. https://doi.org/10.51590/waraqat.v8i2.576

Kim, J., Yu, S., Detrick, R., & Li, N. (2025). Exploring students' perspectives on Generative AI-assisted academic writing. Education and Information Technologies, 30(1), 1265～1300. https://doi.org/10.1007/s10639-024-12878-7

Li, M., Pan, C., Zhang, W., Ling, J., Sun, J., Liu, X., ... & Zhu, J. (2025). Enhancing pre-service teachers training with AI-simulated students: providing realistic classroom experiences without impacting real learners. Interactive Learning Environments, 1-18. https://doi.org/10.1080/10494820.2025.2591238

Liu, Z.-M., Hwang, G.-J., Chen, C.-Q., Chen, X.-D., & Ye, X.-D. (2024). Integrating large language models into

EFL writing instruction: Effects on performance, self-regulated learning strategies, and motivation. Computer Assisted Language Learning, 1～25. https://doi.org/10.1080/09588221.2024.2389923

Morphew, J. (2021). Unskilled and unaware? Examining the metacognitive awareness of low-performing students. Proceedings of the 2021 AERA Annual Meeting. 2021 AERA Annual Meeting. https://doi.org/10.3102/1680899

Muflihah, T., & Authar, N. (2022). The implementation of peer feedback in efl writing class. Konstruktivisme : Jurnal Pendidikan dan Pembelajaran, 13(2), 184～190. https://doi.org/10.35457/konstruk.v13i2.1844

Shen, X., & Teng, M. F. (2024). Three-wave cross-lagged model on the correlations between critical thinking skills, self-directed learning competency and AI-assisted writing. Thinking Skills and Creativity, 52, 101524. https://doi.org/10.1016/j.tsc.2024.101524

Wang, S.-L., & Hwang, G.-J. (2012). The role of collective efficacy, cognitive quality, and task cohesion in computer-supported collaborative learning (Cscl). Computers & Education, 58(2), 679 ～ 687. https://doi.org/10.1016/j.compedu.2011.09.003

Wilby, J. (2022). Motivation, self-regulation, and writing achievement on a university foundation programme: A programme evaluation study. Language Teaching Research, 26(5), 1010 ～ 1033. https://doi.org/10.1177/1362168820917323

Yan, L., Greiff, S., Teuber, Z., & Gašević, D. (2024). Promises and challenges of generative artificial intelligence for human learning. Nature Human Behaviour, 8(10), 1839～1850. https://doi.org/10.1038/s41562-024-02004-5

Yan, Z., & Brown, G. T. L. (2017). A cyclical self-assessment process: Towards a model of how students engage in self-assessment. Assessment & Evaluation in Higher Education, 42(8), 1247 ～ 1262. https://doi.org/10.1080/02602938.2016.1260091

Zhang, J. (2023). Modelling the interplay of writing achievement goals and grit in predicting L2 writing achievements. System, 117, 103118. https://doi.org/10.1016/j.system.2023.103118

Zhu, J., Yang, Y., & Yan, Z. (2024). Relationships between teacher feedback and English writing proficiency in Chinese students: The mediating effect of writing self-regulated learning strategies. System, 123, 103338. https://doi.org/10.1016/j.system.2024.103338